Automation in Construction 68 (2016) 102–113

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

A linked data system framework for sharing construction defect information using ontologies and BIM environments

Do-Yeop Lee [a], Hung-lin Chi [c], Jun Wang [c], Xiangyu Wang [b,c], Chan-Sik Park [a,*]

[a] School of Architecture and Building Science, Chung-Ang University, South Korea
[b] Department of Housing and Interior Design, Kyung Hee University, South Korea
[c] Australasian Joint Research Centre for Building Information Modelling, Curtin University, Australia

ARTICLE INFO

Article history:
Received 19 March 2015
Received in revised form 12 April 2016
Accepted 1 May 2016
Available online xxxx

Keywords:
Defect management
Linked data
Ontology
Building information modeling
SPARQL query

ABSTRACT

Defect data contains knowledge about specific work conditions. In order to prevent reoccurrence of defects, a data feedback mechanism is required. However, most defect data are stored in unstructured ways, resulting in the fundamental problem of data utilization. This paper proposes a novel framework by using BIM and linked data technologies for sharing defect data between heterogeneous data sources in a new way. To demonstrate, a defect ontology is developed, work context information is extracted from BIM models, extracted BIM data is converted to RDF format, and SPARQL queries are implemented. The proposed approach could help BIM software applications to take into account information stemming from the defect management domain. Also, it can reduce data search time and improve the accuracy of search results as well. Therefore, this framework may enable reductions of defect occurrence and improvements in current defect management practices.

1. Introduction

Construction defects are perceived as one of the primary causes of low project productivity, resulting in delays, additional costs, materials, and manpower for defect rectification [1,2]. In order to prevent the reoccurrence of defects, various researches have been conducted, focusing on identifying defect causations [3,4], and analyzing cost–time impacts [5,6]. The results of these research efforts provide valuable statistical information for decision-making, allowing project stakeholders to evaluate their enterprise performance, and identify the factors influencing defect occurrence from a holistic perspective. For instance, 'Poor planning and coordination of resources' was ranked first among site management causes of rework [3] and 'Structural steel' had the most significant impact on the defect cost among subcontractor trades [6]. Such information only provides generalized concepts or terminologies at a high level of abstraction. Therefore, practitioners cannot make a proper decision for defect prevention in practical situations by using statistical information alone without access to actual defect cases that include complex information of specific design/construction tasks. In order to address this issue, it is necessary to look into previous research on data feedback for defect prevention. Palaneeswaran [7] argues that the analysis of defect causation should focus on developing a suitable knowledge feedback mechanism. Chong and Low [8] point out that many defects continue to occur repeatedly as designers fail to obtain important feedback about defects and suggest developing a database system by using existing knowledge. Jang and Seo [9] emphasize the necessity of timely data feedback through a case study on apartment defects. For instance, a window condensation defect had occurred in a previous project and then the same defect reoccurred in their next two similar projects due to the lack of appropriate information feedback. However, there has been little research on the data structure required for implementing actual data feedback systems. Moreover, traditional database systems and keyword-based searches appear to be insufficient to improve information discovery and retrieval [10]. Therefore, an appropriate information platform should be considered to enable users to easily search for individual defect cases and directly access them.

In order to develop this platform, the flow of defect data should be closely linked from the perspective of knowledge management processes. However, the utilization of defect data in practice is very low, due to characteristics such as lack of formal structure in data representations, insufficient data input, only text-based search available, and insufficient sharing of defect data. These characteristics will be discussed in detail in Section 2. The features of defect data could be improved by taking advantages of state-of-the-art technologies. In information systems, ontologies are considered as ideal formal tools to represent domain knowledge, as they allow modeling concepts with semantic relationships. In the context of the Semantic Web, ontologies play an important role for publishing and connecting structured data on the Web known as linked data [11]. In the architecture engineering and construction (AEC) domain, building information models (BIM)

* Corresponding author.
E-mail address: cpark@cau.ac.kr (C.-S. Park).

http://dx.doi.org/10.1016/j.autcon.2016.05.003
0926-5805/© 2016 Elsevier B.V. All rights reserved.

have been utilized as central repositories of building data to facilitate the exchange and interoperability of information in digital format across the project lifecycle. However, within the wider AEC context, BIM models are only single silos of information [12]. In this regard, a defect linked data framework is proposed underlying the ideas developed in the authors' previous paper [13]. The target of this framework is to convert defect data to the ontology-based linked data format for linking and searching defect data between different data sources. To develop this framework, various technical issues are addressed in this paper.

1) Development of a defect ontology for publishing defect data following linked data principles
2) Extraction of work context information from BIM models for generating machine-readable defect cases
3) Conversion of extracted BIM data to the RDF data model
4) Search of relevant defect cases by using SPARQL queries

The proposed approach could help BIM software applications to take into account information stemming from the defect management domain. Also, it could reduce data search time and improve the accuracy of search results as well. Improving the process of utilizing defect information in design and construction phases can prevent the same defect or error occurrence. In addition, this approach can provide a technical solution towards the usage of valuable data stored in public websites or corporate repositories as data silos.

This paper is organized as follows. Section 2 discusses the current obstacles for sharing defect data. In Section 3, related research on BIM, ontology, and linked data is reviewed in order to find applicable solutions to solve the limitations discussed in Section 2. In Section 4, actual defect cases are analyzed to build a defect ontology. Next, a linked data system framework is proposed and tested in Section 5. Finally, the paper concludes by discussing findings and issues to be addressed in future work.

2. Issues of utilizing defect information

Information regarding defects which are identified during a project lifecycle is stored in various ways. However, because of the lack of structure in the defect data and its various characteristics, the data cannot be used for retrieving related information for a given situation during the design and construction phase. This section discusses the problems and issues to be addressed for defect data reusability through comprehensive literature reviews and usage analysis of defect data.

2.1. Lack of formal structure in defect data representations

With the information-intensive nature of the construction industry, large amounts of data are produced throughout project tasks and usually stored in cooperation's knowledge management systems. However, most data are scattered across various company systems and recorded in unstructured formats. Unstructured data is data that does not follow a machine-readable format, so, there is no reliable way to access, analyze, or search for desired information. In addition, it is difficult to generate valuable knowledge through the vast amounts of existing data.

Most defect data are typically available in unstructured formats. The number of defects that occur in a project typically goes in the thousands, including minor and major defects [14]. Such defect data is usually modified and published in defect case books or stored as individual data files for sharing knowledge. However, these files exist as unstructured text-based data, such as PDF or Word documents, spreadsheets, HTML pages, and so on. The contents of defect cases normally consist of the title, description, picture, diagram, defect cause, prevention method, etc. However, the stored contents differ from case to case. Even if defect cases are stored using a machine-readable format, users cannot find and analyze them without an appropriate data collection structure which describes defect case situations in a standard way. Therefore, in order to retrieve and analyze defect data effectively, a comprehensive defect data collection template is needed and should include various types of contents for multiple purposes [13].

First of all, construction context information provides important clues when searching for defect cases similar to the situations that architects and constructors are working on. Context can be defined as 'situations in which a construction entity acts or exists; whereby an entity can be a person involved in construction, an action taken to complete a job, a component built in a project, or a resource utilized to perform the action or build the component' [10]. In the case of defects, the context information that needs to be stored includes construction method, space, component, material, etc. In addition, defect analysis data such as defect root cause, defect type, impact cost and time needs to be input in order to assist further analysis. By combining context and analysis information, it is possible to provide valuable knowledge to project stakeholders both at the specific task level for retrieving relevant data and at the cooperation strategy level for analyzing data. Finally, to easily understand the situation around which a defect occurred, descriptive information from materials such as 2D plans, section drawings, pictures, and specifications is necessary. Therefore, defect data should be recorded using a machine-readable format and should include various types of information for comprehensively explaining defect cases with a standardized structure.

2.2. Insufficient data input

As discussed previously, different types of detailed context information should be recorded for generating structured defect data and improving usability and search accuracy of defect data. When looking at the contents of defect case books, most context information is included in each case. However, this context information is expressed textually in the title and defect description section. Hence, it is only human-readable but not machine-readable. To make the recorded contents machine-readable, defect data collection structures have been proposed by using codes or industry standard classification systems [13,15]. However, entering data manually by following codes (e.g., painting 9915 and internal walls 75) as suggested in [15] or standard classifications (e.g., cement stucco 22-092423 and wall finishes 21-032010) in [13] is labor-intensive and erodes the quality manager's work efficiency.

In current practice, defect data is generated after defects are identified and need to be rectified on-site. Defect data modification only occurs for data sharing in the office. These two phases (defect data rectification and modification) take place over different time frames, they require different information, and they are handled by different parties. First, in the defect identification and rectification process, the required information is commonly a location and identified defect situation, for instance, 'Water leak in the toilet on the 3rd floor'. It is transferred to the sub-contractor by the quality manager who orders defect rectification. At this point, detailed context information is unnecessary because stakeholders involved in a project already have related information through drawings and specifications. Also, defect analysis information is not recorded in this phase because the defect management process focuses on defect identification and rectification rather than data reuse. Next, the process of defect modification is that, after defect rectification work, important cases are selected and then modified for information sharing. If recorded contents are insufficient for the reasons discussed above, additional efforts are required for finding context and analysis data in order to modify and analyze defect cases. If it is impossible, because of a time lag between the two processes, the reusability of defect data would decrease. For that reason, in most research, defect data is manually classified and reformed to enable useful data for research purposes [4,5,8].

In summary, the current defect management process faces a trade-off relationship between task efficiency and the level of data quality. Therefore, to solve this problem, the context information needs to be