PROPOSAL FOR ECO-FRIENDLY CERTIFICATION INFORMATION GENERATION PROCESS USING WEB CRAWLING TECHNOLOGY

[1]SIMHEE HONG, [2]JUNGHO YU

Kwangwoon University
Email: hshgl9786@kw.ac.kr, myazure@kw.ac.kr

Abstract- The application of an eco-friendly certification system is becoming increasingly mandatory in Korea and internationally; as a result, various business improvement processes have been proposed. Studies have been carried out to generate information that can be used for certification work by applying building information modeling (BIM), as well as to review legal codes using the generated information. In addition, studies to establish a classification system for building material database that can be used for the eco-friendly certification system have also conducted. However, there is still a lack of research on systematically managing and continuously updating the generated information. Therefore, in this study, we propose an automated process of generating eco-friendly certification information by building a database using a Web crawler that automatically collects information extracted from the BIM model and unstructured data on the Web. Furthermore, we identify the potential problems that can occur during the Web crawling process as well as how to address them. This study is expected to contribute to the establishment of a database that can reduce the time required for eco-friendly certification and enables the use of accurate information for the certification process through constant updates.

Keywords- Data Crawling, Database, Construction Material, Extracting information

I. INTRODUCTION

Eco-friendly certification has become a mandatory requirement rather than an option. . In the public sector, the application of an eco-friendly certification system is gradually becoming mandatory. Globally, various eco-friendly certification systems, including Leadership in Energy and Environmental Design (LEED) and Building Research Establishment Environmental Assessment Method (BREEAM), have become obligatory, increasing the related work. As a result, efforts have been made in other countries to increase the accuracy of data extraction and to reduce workload by automating the eco-friendly certification process. A study was conducted to automatically extract the necessary information for a LEED system using building information modeling (BIM) [3], and various studies have been carried out to establish a database of building materials that can be used for environmental certification.

The typical eco-friendly certification systems currently implemented in Korea include the health-friendly housing construction codes and Green Standard for Energy and Environmental Design (G-SEED). Healthy eco-friendly housing construction codes were established for the purpose of creating a pleasant indoor environment, while G-SEED consists of eight items related to indoor environment, housing, land, and resources1. The information required for both certification systems relates to the indoor environment, and the required information is largely divided into area, material, and attached documents. Here, if area information generation and material information generation, which are manually created by referring to drawings, are automated, the time required for eco-friendly certification work can be greatly shortened.

As the fourth industrial revolution came into existence, various technologies related to information generation were developed. Among them, Big Data is attracting attention as a technology to collect and analyze unstructured data, which was not possible before, and to provide various information and convenience. Among Big Data technologies, Web crawling technology, which collects and stores data on the Web, can automatically manage the information that needs to be updated periodically. Therefore, a more convenient database can be established if the information on new products that are continuously registered with an expiration period, such as eco-friendly material certification information, can be automatically collected and managed.

Therefore, in this study, we propose the whole process to automatically generate the area and material information required for environment-friendly certification for the BIM model created with standard material names. It also identifies potential problems during Web crawling as well as improvements related to specific sites2 that currently disclose information about eco-friendly materials. This study is expected to contribute not only to the automation of eco-friendly certification but also to the automatic generation of

[1] Land use and transportation, energy and environmental pollution, materials and resources, water circulation management, maintenance, ecological environment, indoor environment, and housing performance.

[2] A homepage that knows HTML structure is defined as a specific site (Chapter 3).

general information, pollutant-related information, and eco-friendly material certification information that can be applied to eco-friendly certification as well as building area information.

II. LITERATURE REVIEW

Various studies have been conducted in Korea on the automation of information generation. First, on the automatic generation of area information, studies were continuously carried out to extract surface information, line information, and coordinate information by extracting the Industry Foundation Classes (IFC) files included in the BIM model. The BIM object to be extracted was first specified, then the IFC file included in the object was extracted for value generation, and information accuracy was improved by inserting a review algorithm inserting a review algorithm for value generation. In addition, studies on how to use the information extracted from BIM for high-rise evacuation regulations [4] and the review of laws for the disabled [5] were conducted in the automatic review of legal standards. Therefore, if the extracted information is used the same way, it will be possible to develop a process that can automatically judge whether the eco-friendly certification standard is suitable using the extracted value.

Regarding the automation of material information generation, research on database construction has been steadily carried out to manage information effectively. Studies have also been conducted to develop a material classification system that can be utilized properly. Among the applicable construction materials and related product information, the information demanded by a large number of users have been classified and assigned with ID codes accordingly. The status of data management relating to domestic and foreign materials has been analyzed to establish a classification system for each material and the necessity of building an eco-friendly material database was suggested [1]. A database with subdivisions was constructed containing pollutant emission data by comparing and analyzing the amount of pollutants emitted by each material [2]. Therefore, the information generation process can be automated if the material database used in the eco-friendly certification system is constructed and the process for continuous data management is provided.

III. OVERALL PROCESS TO CREATE ECO-FRIENDLY CERTIFICATION INFORMATION

The extraction process for area information and material information required for eco-friendly certification consists of two parts (Fig. 1). First, area-related information is extracted from the BIM model. The extracted information is then further divided into width information and length information. If the area-related information is extracted, it is necessary to confirm whether the value of the designated area was correctly extracted according to the distance between members. In Revit, which is a tool typically used for BIM, a value is automatically extracted based on the centerline, and therefore an algorithm for checking the extracted position must be added to complete the data extraction.

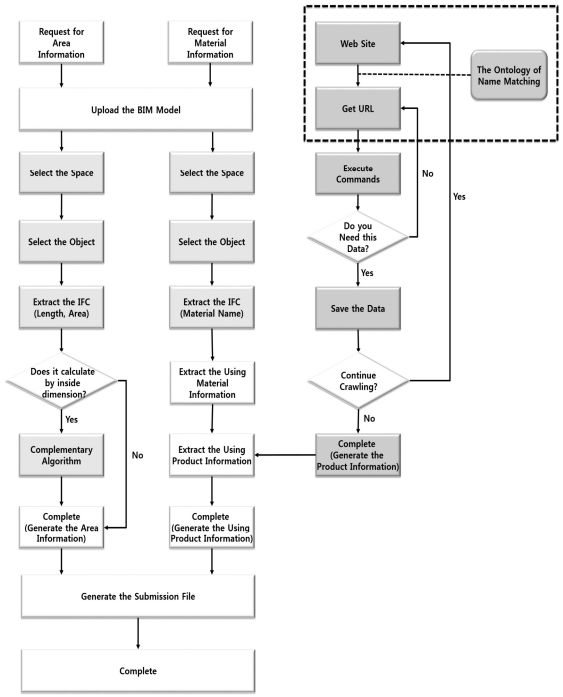


Fig. 1 Data extraction performance map for information generation for areas and materials

Next, material-related information is extracted through the external information storage database, which is stored by crawling Web data. The home page where the extraction is executed is divided into a site whose Hypertext Markup Language (HTML) structure of the home page is known and a site whose HTML structure is unknown, from which information then needs to be extracted using a different method from the existing crawler, such as finding the homepage Uniform Resource Locator (URL) through an additional process. Therefore, Web crawling is generally performed on a homepage whose HTML structure is known, which is defined as a specific site. Information extraction performed at a specific site is executed as it grasps the HTML structure and executes the commands depending on the structure (Fig. 2).

If a Web crawl is performed on a specific site, the HTML structure of the page first needs to be understood. The code must include a command to be executed in the HTML structure requiring command execution and the information extraction process is performed accordingly. The crawl is completed by storing the necessary information according to the defined table format from the extracted information.

Here, there are two kinds of problems that can occur. First, if the number of specific sites increases, the configuration format in which the information is stored according to the homepage may be changed. This problem can be solved by defining the table that cleans up the data and organizing the necessary information according to the prescribed format. Second, the terms contained in a particular site may be different. This problem can be solved by defining the standard material name and applying the ontology technique based on the defined name. For example, if a material name defined by a standard material name is "gypsum board," the name of the material extracted by "gypsum board", "gypsum" "board", and "Gypsum Board" should be converted into the standard material name, and the information about the extracted product should be defined accordingly.

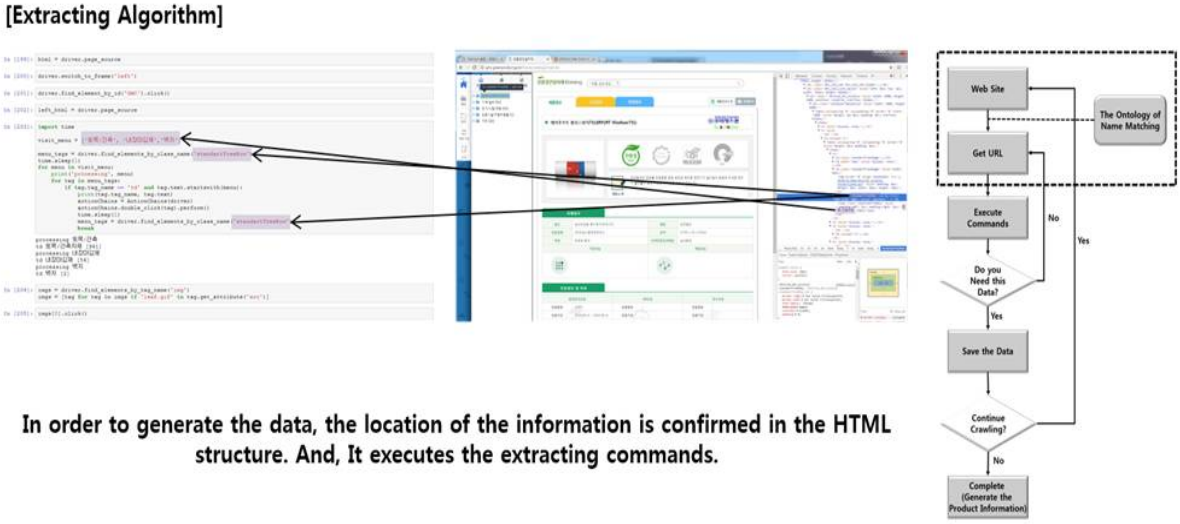


Fig. 2 Data extraction process from a page with a known structure

CONCLUSION

The application of an eco-friendly certification system and the use of eco-friendly materials are increasingly attracting attention, and studies have been continuously carried out to improve processes for more convenient execution. In this paper, we proposed a process to automate information generation by an eco-friendly certification system as well as a method to automatically manage eco-friendly materials using Web crawling. Moreover, in order to construct the eco-friendly materials database, potential problems that may arise because of the Web environment while collecting product information are summarized, and respective improvement measures were suggested. Future studies will test the effectiveness of the improvements to resolve these problems and apply the entire process to an actual project for verification. This study is expected to contribute to basic research that can shorten the time required for eco-friendly certification and can generate and manage related information more conveniently.

ACKNOWLEDGEMENTS

This research was supported by a grant(16CTAP-C114926-01) from Infrastructure and transportation technology promotion Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

REFERENCES

[1] D. H. Kang, M. S. Yeo, K. W. Kim, “The Method of Classification of Eco-Friendly Material and Establishing a Database”, Magazine of the Society of Air-Conditioning and Refeigerating Engineers of Korea(SAREK), 35(12), 33-40, 2006

[2] Y. J. Yu, C. W. Lee, M. G. Kim, “Development of a Building materials database; Volatile organic compounds, formaldehyde emission rates and chemical compositions”, Analytical Science & Technology, 22(1), 57-64, 2009

[3] A. Salman, A. C. Wade, O. Darren, A. Irtishad, “Building information modeling for sustainable design and LEED rating analysis”, Automation in Construction, 20, pp.217-224, 2010

[4] G. H. Cho, “Development of Automated Checking Modules for Open BIM-based Quality Assurance in Architecture Design Phase(Focused on the Applying Fire Evacuation Regulations for Supertall Building)”, MS thesis, KyungHee Univ., Korea, 2012

[5] J. S. Kang, “A Study on the Development of Architectural Administration for Code Checking Support System in Open BIM Environments (Focused on Technology of the BIM Data Management)”, Doctor of Philosophy, KyungHee Univ., Korea, 2013

[6] D. M. Seo, H. M. Jung,”Intelligent Web Crawler for Supporting Big Data Analysis Services”, The Journal of the Korea Contents Society, 13(12), 575-584, 2013

[7] Korea Ministry of Government Legislation (2016).

[8] Korea Environment Industry & Technology Institute(KEITI), Green Construction Material Information system , <http://gmc.greenproduct.go.kr/main.do> (Jan. 1, 2017)

★★★